Best Available Copy

Final Report on Contract N00014-85-K-0445 on Machine Learning

L.G. Valiant
Harvard University
Cambridge, MA 02138

Work was centered on quantifiable aspects of machine learning, particularly efficient algorithms for learning new classes of representations, and proofs of limitations. We list here the publications arising from the last four years of the work, omitting earlier work.

1. B. Awerbuch, I. Cidon, S. Kutten and Y. Mansour and D. Peleg. Broadcast with Partial Knowledge. *Proc. ACM Symposium on Principles of Distributed Computing* (1991)

2. A. Blum, L. Hellerstein and N. Littlestone. Learning in the presence of finitely or infinitely many irrelevant attributes. *Proc. 4th COLT* (1991) 157-167.

3. A. Ehrenfeucht, D. Haussler, M. Kearns and L.G. Valiant. First Workshop on Computational Learning Theory, MIT, Aug 3-5 (1988) 110-120.

4. S. Even and Y. Mansour. A construction of a pseudorandom cipher from a single pseudorandom permutation. Submitted for publication.

5. Q. Gao and M. Li. The minimum description length principle and its application to on line learning of handprinted characters. Proc. IJCAI-89, (1989) 843-848.

6. M. Geréb-Graus and M. Li. Three one-way heads cannot do string matching. *J. Comp. Syst. Sci.* (to appear).

7. T. Hancock. Identifying μ-Formula Decision Trees with Queries, Proc. COLT 90, 23-37.

8. T.R. Hancock. Learning 2μDNF formulas and $k\mu$ decision trees. *Proc. 4th COLT* (1991) 199-212.

9. T.R. Hancock and L. Hellerstein. Learning read-once formulas over fields and extended bases. *Proc. 4th COLT* (1991) 326-336.

10. T.R. Hancock and Y. Mansour. Learning monotone $k\mu$ DNF formulae on product distributions. *Proc. 4th COLT* (1991) 179-183.

11. T.R. Hancock. Identifiability is closed under embeddings in read-once formulas and μ-decision trees. Harvard University, Technical Report TR-16-91, Aiken Lab. (1991).

12. T. R. Hancock, N.H. Bshouty, and L. Hellerstein. Learning Boolean Read-Once Formulas over Generalized Bases. To appear in *JCSS*.

13. T. R. Hancock, M. Golea, and M. Marchand. Learning Nonoverlapping Perceptron Networks From Examples and Membership Queries. To appear in *Machine Learning*.

14. T. R. Hancock. Learning $k\mu$ Decision Trees on the Uniform Distribution. To appear in *COLT* 1993.

15. T. R. Hancock, N. H. Bshouty, S. A. Goldman, and S. Matar. Asking Questions to Minimize Errors. *COLT* 1993.

16. D. Helmbold, R. Sloan and M.K. Warmuth. Learning Nested Differences of Intersection-closed Concept Classes, Machine Learning *5* (1990) 165-196.

17. D. Helmbold, R. Sloan and M.K. Warmuth. Learning Integer Lattices. COLT 90, 288-302.

18. A. Israeli and M. Li. Bounded Time-Stamps. Proc. 28th IEEE Symp. on Foundations of Computer, Los Angeles, CA, May 28-29, (1987) 371-382.

19. M. Kearns and M. Li. Learning in the Presence of Malicious Errors. Proc. 20th ACM Symp. on Theory of Computing. Chicago, Illinois, May 2-4 (1988) 267-280.

20. M. Kearns and L.G. Valiant. Cryptographic limitations on learning Boolean formulae and finite automata. Proc. 21st ACM Symp. on Theory of Computing, New York, NY (1989) 433-444.

21. M. Kearns and L. Pitt. A polynomial-time algorithm for learning k-variable pattern languages from examples. Proc 2nd Workshop on Computational Learning Theory. Morgan Kaufmann, San Mateo, CA (1989) 57-71.

22. M. Kearns. The Computational Complexity of Machine Learning, MIT Press, Cambridge, MA (1990).

23. M. Kearns, M. Li and L.G. Valiant. Learning Boolean Formulae. JACM (to appear).

24. M. Kearns and L.G. Valiant. Cryptographic limitations to learning Boolean formulae and finite automata JACM (to appear.)

25. E. Kushilevitz and Y. Mansour. Learning decision trees using the Fourier spectrum. *Proc. 23rd ACM Symp. on Theory of Computing* (1991) 455-464.

26. M. Li and P. Vitanyi. Two Decades of Kolmogorov Complexity. TR-02-88, Aiken Laboratory, Harvard University, 1988.

27. M. Li and P. Vitanyi. A Very Simple Construction for Atomic Multi-writer Register. TR-01-87, Aiken Laboratory, Harvard, 1987.

28. M. Li and Y. Yesha. New lower bounds for parallel computation. J. of ACM 36:3 (1989) 671-680.

29. M. Li and P. Vitanyi. How to share concurrent asynchronous wait-free variables. Proc. 16th International Coll. on Automata, Languages and Programming, Springer Lecture Notes in Computer Science, 372 (1989) 488-505.

30. M. Li and P. Vitanyi. Inductive Reasoning and Kolmogorov Complexity. *J. Comp. Syst. Sci.* (to appear).

31. M. Li and P. Vitanyi. Kolmogorov complexity and its applications. In *Handbook of Theoretical Computer Science*, (J. van Leeuwen, ed.), North Holland, (1991) 187-254.

32. M. Li and Y. Yesha. Resource bounds for parallel computation of threshold and symmetric functions. *J. Comp. Sys. Sci. 42* (1991) 119-137.

33. N. Littlestone. From on-line to batch learning. Proc. 2nd Workshop on Computational Learning Theory. Morgan Kaufmann, San Mateo, CA (1989) 269-284.

34. N. Littlestone and M.K. Warmuth. The Weighted Majority Algorithm. Proc. 30th IEEE Symp. on Foundation of Computer Science, (1989) 256-267.

35. N. Littlestone. Notes on the derivation and quality of Chernoff-type bounds for sums of bounded random variables. Manuscript.

36. N. Littlestone. Redundant noisy attributes, attribute errors, and linear-threshold learning using winnow. *Proc. 4th COLT* (1991) 147-156)

37. N. Littlestone, P. Long and M. Warmuth. On-line learning of linear functions. *Proc. 23rd ACM Symp. on Theory of Computing* (1991) 465-475.

38. Y. Mansour and B. Patt-Shamir. Greedy packet scheduling on shortest paths. *Proc. ACM Symposium on Principles of Distributed Computing* (1991).

39. L. Pitt and C.H. Smith. Probability and Plurality for aggregations of Learning Machines. Inf. and Compt. *77* (1988) 77-92.

40. L. Pitt and L.G. Valiant. Computational limitations on learning from examples. J. of ACM 35:4 (1988) 965-985.

41. L. Pitt and M.K. Warmuth. The minimum consistent DFA problem cannot be approximated within any polynomial. Proc. 21st ACM Symp. on Theory of Computing, New York NY (1989) 421-432.

42. L. Pitt and M.K. Warmuth. Prediction preserving reducibility. *J. of Comp. and Syst. Sci. 41* (1990) 430-467.

43. D. Ratner and M.K. Warmuth. The $(n^2 - 1)$-puzzle and related relocation problems. J. of Symbolic Computation. (to appear).

44. L.G. Valiant. Functionality in Neural Nets. Proc. 7th National AAAI Conference on Artificial Intelligence, Morgan Kauffman, San Mateo, CA (1988).

45. L.G. Valiant. Optimally universal parallel computers. Phil. Trans. R. Soc. London, A326 (1988) 373-376.

46. L.G. Valiant. Why is Boolean Complexity Theory is Difficult? *Boolean Function Complexity,* (M.S. Paterson ed.), Lond. Math. Soc. Lecture Note Series *169* (1992) 84-94.

47. L.G. Valiant. General purpose parallel architectures. In *Handbook of Theoretical Computer Science* (J. van Leeuwen, ed.), North Holland, Amsterdam (1990) pp. 944-971.

48. L.G. Valiant. A view of computational learning theory. In *Computation and Cognition* (C.W. Gear, ed.), Soc. Ind. and Appl. Math., Philadelphia, (1990) 32-53.

We shall mention five highlights from the above papers:

(a) NP-completeness results: In [42] it was shown that some very restricted representation classes, such as 2-term dnf, are difficult to learn, if the learner is forced to express the hypothesis in the restricted class, but becomes tractable if the learner is allowed a more expressive representation in learning. The technique has been applied by others since to other representations, such as neural nets.

(b) Representation independent hardness results: In [20] it was shown that general Boolean formulas and finite automata are hard to learn in the pac model (assuming some cryptographic conjectures) whatever representation is chosen by the learner. This says, roughly, that a black box containing an unknown formula can behave essentially as a random function, about which little can be found out by feasible experimentation. Some remarkable extensions of these results have been developed recently by others.

(c) Algorithms for learning several new subclasses of decision trees and disjunctive normal form were discovered by Hancock and described in his thesis and several papers. A powerful application of the Fourier spectrum method was applied by Kushilevitz and Mansour to learn decision trees under the uniform distribution with membership queries [27].

(d) The concept of weak learning was introduced [20, 25]. The result of Schapire while at MIT and others subsequently showed that any weak learning algorithm in the pac model could be "boosted" to higher accuracy by a very general procedure. Recent experiments at AT&T suggest that this boosting method is applicable to natural data, such as images of digits.

(e) A monograph based on the neuroidal model introduced in [44] is near completion. The model is an attempt at a basis for giving a computational account of some of the most basic tasks of memorization and learning. The tasks considered are restricted to "random access" tasks, which are defined to be those that potentially involve any part of memory. In this context the model aims to capture the relevant properties of real neurons in cortex. The algorithms developed attempt to be consistent with the quantitative parameters of human performance, such as processing time and memory capacity.